The EMBO Journal vol.14 no.17 pp.4336–4349, 1995

Identification and characterization of three members of the human SR family of pre-mRNA splicing factors

Gavin R.Screaton[1], Javier F.Cáceres[2], Akila Mayeda[2], Martyn V.Bell, Magda Plebanski, David G.Jackson, John I.Bell and Adrian R.Krainer[2]

Molecular Immunology Group, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK and [2]Cold Spring Harbor Laboratory, PO Box 100, 1 Bungtown Road, Cold Spring Harbor, NY 11724-2208, USA

[1]Corresponding author

SR proteins have a characteristic C-terminal Ser/Arg-rich repeat (RS domain) of variable length and constitute a family of highly conserved nuclear phosphoproteins that can function as both essential and alternative pre-mRNA splicing factors. We have cloned a cDNA encoding a novel human SR protein designated SRp30c, which has an unusually short RS domain. We also cloned cDNAs encoding the human homologues of *Drosophila* SRp55/B52 and rat SRp40/HRS. Recombinant proteins expressed from these cDNAs are active in constitutive splicing, as shown by their ability to complement a HeLa cell S100 extract deficient in SR proteins. Additional cDNA clones reflect extensive alternative splicing of SRp40 and SRp55 pre-mRNAs. The predicted protein isoforms lack the C-terminal RS domain and might be involved in feedback regulatory loops. The ability of human SRp30c, SRp40 and SRp55 to modulate alternative splicing *in vivo* was compared with that of other SR proteins using a transient cotransfection assay. The overexpression of individual SR proteins in HeLa cells affected the choice of alternative 5′ splice sites of adenovirus E1A and/or human β-thalassemia reporters. The resulting splicing patterns were characteristic for each SR protein. Consistent with the postulated importance of SR proteins in alternative splicing *in vivo*, we demonstrate complex changes in the levels of mRNAs encoding the above SR proteins upon T cell activation, concomitant with changes in the expression of alternatively spliced isoforms of CD44 and CD45.

Keywords: alternative splicing/RNA splicing/SR proteins

Introduction

Pre-mRNA splicing requires numerous protein factors that associate with U1, U2, U4, U5 or U6 snRNAs and/or with the pre-mRNA substrate during spliceosome assembly (reviewed in Lamm and Lamond, 1993; Moore *et al.*, 1993). A number of cDNAs encoding essential metazoan splicing factors have been isolated. Some, but not all, of these factors have one or more copies of the RNA recognition motif (RRM), which is common among RNA binding proteins and includes the conserved RNP-1 and RNP-2 submotifs (reviewed in Bandziulis *et al.*, 1989; Kenan *et al.*, 1991; Birney *et al.*, 1993). In addition, several of these splicing factors contain regions rich in Arg and Ser residues (RS domains). Some of these factors, which are members of the SR protein family, have C-terminal RS domains of variable length with a characteristic repeat periodicity. Somewhat different RS domains are found at different positions in other splicing factors from several species, including the U1-70K polypeptide, both subunits of the splicing factor U2AF and the *Drosophila* splicing regulators Tra, Tra2 and SWAP (reviewed in Birney *et al.*, 1993).

The SR protein family is remarkably conserved, with at least five polypeptides of characteristic electrophoretic mobilities (~20, 30, 40, 55 and 70–75 kDa apparent molecular weights) found in cells from a variety of species (Zahler *et al.*, 1992). Human cDNAs encoding five SR proteins have been cloned: SF2/ASF (SRp30a; Ge *et al.*, 1991; Krainer *et al.*, 1991) and its two alternatively spliced isoforms designated ASF-2 and ASF-3 (Ge *et al.*, 1991), SC35 (SRp30b or PR264; Fu and Maniatis, 1992b; Vellard *et al.*, 1992), SRp20 (X16 in the mouse and RBP1 in *Drosophila*; Ayane *et al.*, 1991; Kim *et al.*, 1992; Zahler *et al.*, 1992), SRp75 (Zahler *et al.*, 1993b) and 9G8 (Cavaloc *et al.*, 1994). HRS is the rat homologue of SRp40 (Birney *et al.*, 1993; Diamond *et al.*, 1993). cDNAs encoding two closely related variants of SRp55, one of which was named B52, were isolated from *Drosophila* (Champlin *et al.*, 1991; Roth *et al.*, 1991), and homologous cDNAs encoding several SR proteins from other species have been described (see Birney *et al.*, 1993).

SR proteins function as both essential and alternative splicing factors *in vitro* and/or *in vivo* (Ge *et al.*, 1991; Krainer *et al.*, 1991; Fu *et al.*, 1992; Mayeda *et al.*, 1992; Zahler *et al.*, 1992, 1993a; Cáceres *et al.*, 1994). For some, if not all, SR proteins, alternative splicing activity can be antagonized in a dose-dependent manner by hnRNP A1 (Fu *et al.*, 1992; Mayeda and Krainer, 1992; Mayeda *et al.*, 1993) and, in addition, by hnRNP $A1^B$, A2 and B1 to varying extents (Mayeda *et al.*, 1994). Sequence-specific interactions have been characterized between SR proteins and certain 5′ or 3′ splice site sequences (Fu and Maniatis, 1992a; Staknis and Reed, 1994; Zuo and Manley, 1994), as well as exonic splicing enhancer elements (Lavigueur *et al.*, 1993; Sun *et al.*, 1993; Tian and Maniatis, 1993; Staknis and Reed, 1994). The binding of SR proteins to pre-mRNA appears to be one of the earliest steps in spliceosome assembly, enhancing the binding of U1 snRNP to the 5′ splice site to form a commitment complex (Eperon *et al.*, 1993; Fu, 1993; Kohtz *et al.*, 1994; Staknis and Reed, 1994). Based on the observation that SR proteins can interact with U1-70K and $U2AF^{35}$ (both of which contain RS domains), it has also been postulated that SR

proteins enhance splicing by bringing the 5′ and 3′ splice sites into closer proximity via a bridge of protein contacts (Wu and Maniatis, 1993; Kohtz *et al.*, 1994). This is consistent with the formation of similar complexes containing SR proteins with pre-mRNA fragments containing only 5′ or 3′ splice sites or enhancer elements (Staknis and Reed, 1994).

Considerable overlap in SR protein function is demonstrated by their ability individually to complement an S100 extract deficient in all SR proteins (Krainer *et al.*, 1990; Fu *et al.*, 1992; Kim *et al.*, 1992; Mayeda *et al.*, 1992; Zahler *et al.*, 1992). It is unlikely, however, that individual members of the family are functionally redundant, as relative differences in their specific activities in constitutive or alternative splicing with a variety of substrate pre-mRNAs have been described (Kim *et al.*, 1992; Fu, 1993; Sun *et al.*, 1993; Zahler *et al.*, 1993a). Moreover, *Drosophila* SRp55/B52 has been shown to be an essential gene (Ring and Lis, 1994; S.Mount, personal communication), and the strong phylogenetic conservation of the entire set of SR proteins with minimal or no sequence variation between homologous family members in different species (Birney *et al.*, 1993) provides a forceful argument against functional redundancy.

A detailed mutational analysis of SF2/ASF (Cáceres and Krainer, 1993; Zuo and Manley, 1993) has demonstrated the requirement for the RS domain, including both Arg and Ser residues, in constitutive splicing. In the presence of limiting amounts of wild-type SR proteins to support basal splicing, SF2/ASF with a mutated or deleted RS domain can still promote alternative splice site switching in a concentration-dependent manner (Cáceres and Krainer, 1993; Zuo and Manley, 1993). In addition, both RRMs in SF2/ASF are required for constitutive and alternative splicing *in vitro*, despite the fact that some SR proteins lack the second RRM.

We have isolated three cDNAs encoding SR proteins that show extensive sequence homology to other members of the SR protein family. The new sequences include the human homologues of *Drosophila* SRp55/B52 and rat SRp40/HRS, as well as a novel SR protein, which we designate SRp30c. Additional cDNA clones represent alternatively spliced isoforms of SRp55 and SRp40. We show that recombinant human SRp30c, SRp55 and SRp40 function individually as essential splicing factors *in vitro*. We have also studied the effect of overexpression of these and other SR proteins on alternative splicing *in vivo*, and report differences in their effects on the alternative splicing of adenovirus E1A pre-mRNA. Consistent with the potential importance of these proteins in the control of tissue-specific alternative splicing, we have observed complex changes in SR protein gene expression following T cell activation.

Results

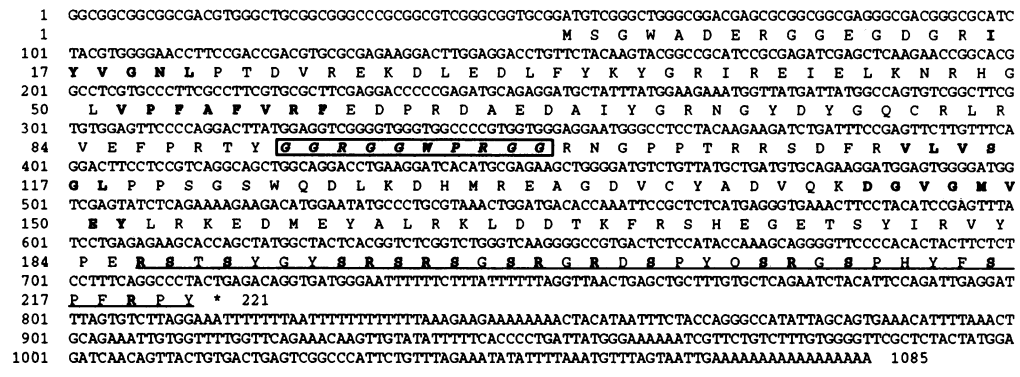
Cloning of novel human cDNAs encoding SR proteins

A sequence analysis of human SF2/ASF and SRp75, as well as *Drosophila* SRp55/B52, reveals extensive overall homology and local regions of identity in the N-terminal RRM and the central atypical RRM (Birney *et al.*, 1993). The second RRM in these proteins contains the invariant signature SWQDLKD, although other SR proteins, such as SC35, SRp20 and the more recently identified 9G8, lack the atypical RRM. We used PCR with degenerate primers to isolate probes for new SR proteins based on amplification between the above heptapeptide region and the conserved RNP-1 submotif region of the N-terminal RRM. PCR products of the expected size were generated using a mixture of cDNAs made by oligo(dT) priming of RNA extracted from three human cell lines: HT29 (colon adenocarcinoma), HeLa (cervical carcinoma) and MOLT-4 (T cell leukaemia). Nucleotide sequencing of cloned PCR products revealed three novel cDNA sequences with homology to SR proteins in the amplified region between the degenerate primers (data not shown).

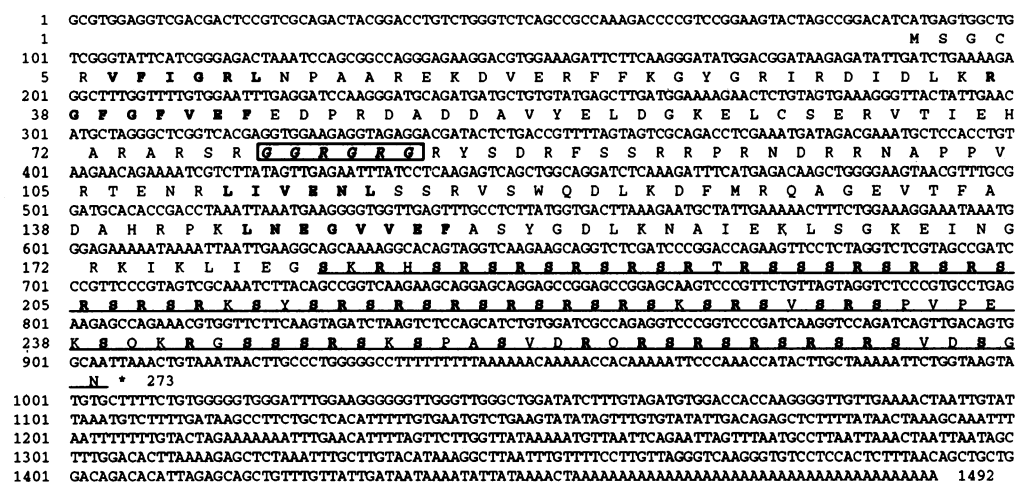
Oligonucleotide probes based on these unique sequences were used to isolate full-length cDNA clones from an HT29 cDNA library. These clones were designated SRp30c, SRp40 and SRp55, and they encode proteins with predicted molecular weights of 25.5, 31.2 and 39.6 kDa, respectively. The usual nomenclature for SR proteins is based on approximate apparent molecular weights on SDS–PAGE (Zahler *et al.*, 1992), although the proteins typically migrate more slowly than expected from their predicted molecular weights, in part due to phosphorylation (see below). The three nucleotide and predicted amino acid sequences are given in Figure 1. All three open reading frames have the sequence features expected for SR proteins, including an N-terminal RRM, a glycine-rich 'hinge', a second RRM with the invariant SWQDLKD signature and a C-terminal RS domain that varies greatly in length among the three predicted proteins. The open reading frames of the human SRp40 and SRp55 cDNAs include amino acid sequences that are identical to the reported sequences of several short peptides derived from purified HeLa cell SRp40 and SRp55 (Zahler *et al.*, 1992). SRp55 is similar in length to its *Drosophila* homologue SRp55/B52 (344 versus 350 or 376 amino acids), and is 64% identical to it in amino acid sequence. Human SRp55 is 65% identical to SRp40 and 84% identical to SRp75. Human SRp40 is 99% identical at the amino acid level and 94% identical at the nucleotide level within the coding sequence to rat HRS, an insulin-regulated gene (Diamond *et al.*, 1993) that was previously proposed to be rat SRp40 (Birney *et al.*, 1993). The third cDNA encodes a predicted protein that is very similar in size (25.5 kDa) to SF2/ASF (SRp30a; 27.7 kDa), SC35 (SRp30b; 25.6 kDa) and 9G8 (27.4 kDa), but is clearly a distinct SR family member which we have named SRp30c.

The expression of SRp30c was analysed by Northern blot hybridization of total RNA from selected mouse tissues (Figure 2A). Cloning and preliminary sequence analysis revealed the existence of a highly conserved mouse homologue of human SRp30c (data not shown). Hence, as expected, the human probe hybridized specifically to an RNA of ~1.1 kb. The integrity of the RNA from each tissue was verified by reprobing the same filter with a mouse β-actin probe (Figure 2B). The expression of SRp30c mRNA varied considerably and was highest in kidney, lung and spleen. Intermediate levels were detected in liver and small intestine, whereas much lower levels were present in brain, heart and skeletal muscle. In addition, the expression of SRp30c was detected by RT-PCR in HeLa, HT29 and MOLT-4 human cell lines (data

SRp30c



SRp40



SRp55

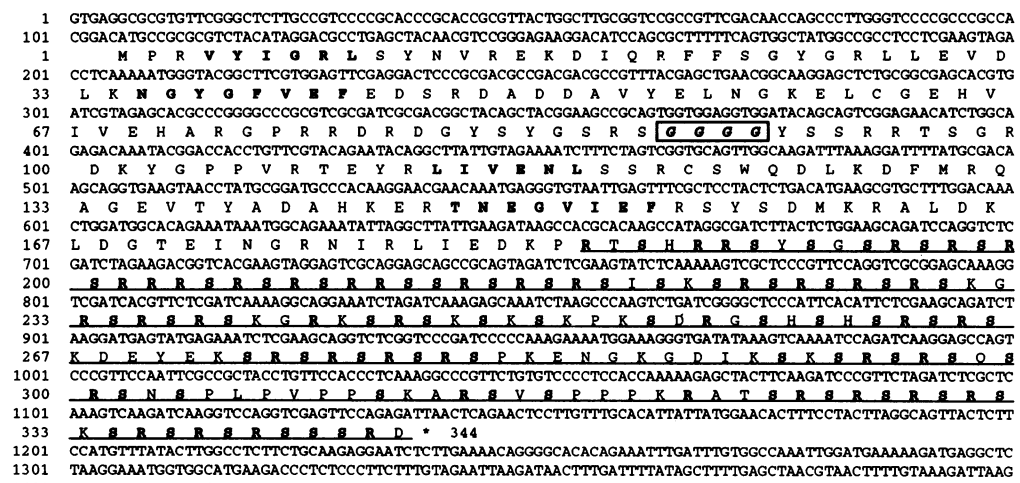


Fig. 1. Nucleotide and deduced amino acid sequences of SRp30c-1, SRp40-1 and SRp55-1 cDNAs. RNP-2 and RNP-1 submotifs are shown in bold, RS domains are underlined, with Arg and Ser residues further highlighted in bold, and glycine-rich tracts separating the two RRMs shown as boxed italics. Using SRp40 as an example, the degenerate primers RRM1b (DADDAVYE) and RRM2b (WQDLKDY/HM) will anneal to DNA sequences corresponding to amino acids 49–56 and 121–128, respectively. GenBank accession numbers: U30825 (SRp30c), U30826 (SRp40-1) and U30883 (SRp55-1).

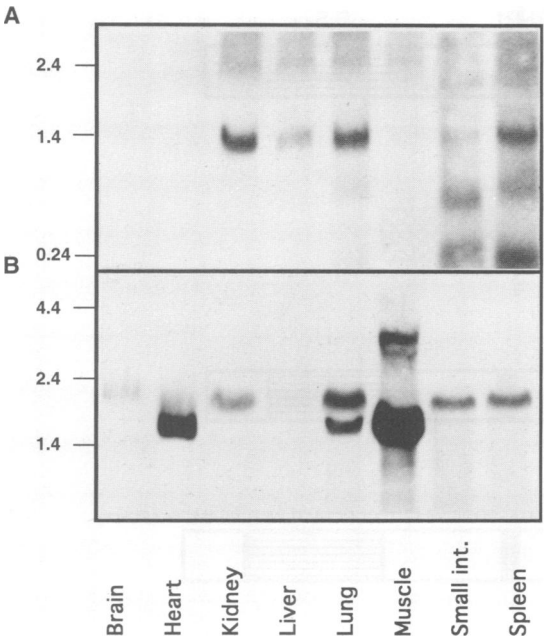


Fig. 2. Expression of SRp30c in adult mouse tissues. Samples of 40 μg of total RNA from the indicated adult mouse tissues were analysed by Northern blot hybridization with a SRp30c-specific probe (**A**), as described in Materials and methods. A transcript of ~1.1 kb is observed. The filter was stripped and reprobed with a mouse β-actin probe as a control (**B**). The relative mobilities of RNA size markers are shown on the left.

not shown). The tissue-specific expression of SRp75, SRp55, SRp40, SF2/ASF, SC35 and SRp20 from various species has been demonstrated previously (reviewed in Birney *et al.*, 1993). Such variations in the expression of individual SR proteins may contribute to the tissue-specific expression of alternatively spliced isoforms of numerous cellular genes.

Alternative splicing generates SRp55 and SRp40 isoforms

We sequenced several independent cDNA clones of SRp55 (five clones analysed) and SRp40 (three clones analysed). There were multiple differences between these cDNA clones, which most probably reflect the alternative splicing of SRp55 and SRp40 pre-mRNAs (Figure 3). SRp40-1 and SRp55-1 cDNAs are considered to encode the bona fide SRp40 and SRp55 proteins, because they have all the domains that are conserved in SR proteins that have demonstrable biochemical functions. All three SRp40 cDNA clones differed from each other, reflecting alternative splicing at four sites, including one in the 5′ untranslated region (UTR). Clone SRp40-2 differs from SRp40-1 in three places. The first difference is a 128 bp insertion in SRp40-2 before the initiation codon. The second difference is an insertion of 285 bp between RRM1 and RRM2, which appears to result from the retention of an intron, as the insert is bordered by sequences that match the 5′ and 3′ splice site consensus sequences. Translation through this insert would lead to premature termination before the second RRM. The third difference in the SRp40-2 cDNA clone is a switch to an alternative 3′ sequence which, if this were the only change, would lead to premature termination before the RNP-1 motif of RRM2. Clone SRp40-3 also has the 128 bp and 285 bp insertions and, in addition, lacks 3 bp corresponding to Arg185 in SRp40-1. The missing sequence is TAG, which overlaps codons 184 and 185 (Figure 1) and may be recognized as an alternative 3′ splice site or may represent an allelic difference.

Two putative alternatively spliced isoforms of SRp55 were represented by single cDNA clones. The SRp55-2 cDNA contains an insert of 268 bp, which places nonsense codons between RRM1 and RRM2 and would generate a shorter protein with a different 50 amino acid C-terminus. The SRp55-3 cDNA encodes a protein that has a different sequence in place of the terminal 32 amino acids of the RS domain. However, because the 3′ end of the mRNA is missing from this cDNA clone, the structure of the C-terminus of SRp55-3 is unknown.

Expression of recombinant SR proteins

The cDNAs encoding human SRp55, SRp40 and SRp30c were individually subcloned into the *Escherichia coli* expression vector pET9c, along with cDNAs encoding the previously described human SRp20, SC35 and SRp75 proteins. SR proteins are known to be extensively phosphorylated, leading to an increase in the apparent molecular weight on SDS–PAGE (Ge *et al.*, 1991; Krainer *et al.*, 1991; Fu *et al.*, 1992; Zahler *et al.*, 1992; Gui *et al.*,. 1994). When the cDNAs were analysed by coupled transcription and translation with T7 RNA polymerase in a rabbit reticulocyte lysate, in which partial or complete phosphorylation appears to occur, the observed electrophoretic mobilities of SRp30c, SRp40 and SRp55 corresponded to apparent molecular weights of 27, 37 and 51 kDa, respectively (Figure 4A, lanes 4–6). The cDNA clones encoding SRp20, SF2/ASF, SC35 and SRp75 also resulted in proteins of approximately the expected sizes (Figure 4A, lanes 1–3 and 7). These results confirm the identity and integrity of the SRp55 and SRp40 cDNAs, and suggest that modified SRp30c is likely to comigrate with SF2/ASF, SC35 and 9G8 as part of a broad set of bands (of 30–35 kDa) that is usually observed with SR proteins from mammalian sources (Zahler *et al.*, 1992).

When the same cDNAs were expressed in *E.coli*, where the phosphorylation of SR proteins does not take place, SRp30c, SRp40 and SRp55 had apparent molecular weights of 27, 34 and 32 kDa, respectively (Figure 4B). SRp55 migrated considerably faster than expected from its predicted molecular weight of 39.6 kDa (lane 3). The recombinant proteins reacted with an antibody that recognizes unphosphorylated RS domains (data not shown), suggesting that the unexpected small size of SRp55 expressed in *E.coli* is not due to premature translation termination prior to the C-terminal RS domain or to extensive degradation of the C-terminus. However, it is likely that recombinant SRp55 is missing a portion of the C-terminal RS domain, although the recombinant protein was found to be active in splicing (see below). SRp30c and SRp40 were expressed at high levels in *E.coli*, and SRp55 was expressed at moderate levels. For unknown reasons, other SR proteins, such as SRp20, SC35, SRp75 (data not shown) and 9G8 (Cavaloc *et al.*, 1994), are expressed extremely poorly in *E.coli*.

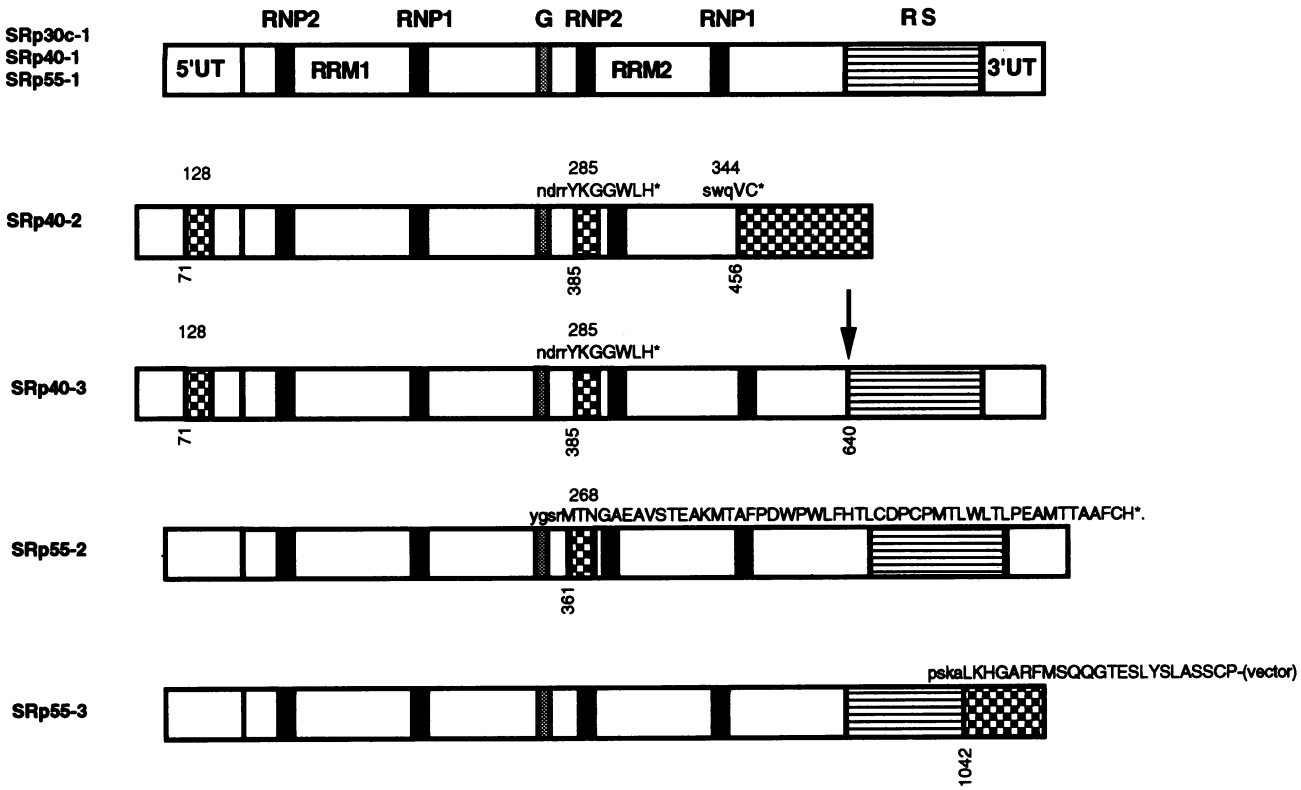


Fig. 3. Structure of cDNAs encoding putative alternatively spliced isoforms of SRp40 and SRp55. A graphic representation of the domains and motifs common to bona fide SRp30c, SRp40 and SRp55 is shown at the top. The RS domain and the 5′ and 3′ UTRs are of arbitrary length. The two RRMs with their respective RNP-2 and RNP-1 submotifs, and the Gly-rich hinge region, are shown. Sequence inserts and changes in the alternatively spliced isoforms of SRp40 and SRp55 are shown as checkered segments. The point of insertion relative to the nucleotide sequences in Figure 1 is given below as a vertical number. The length (bp) of the insert is given above as a horizontal number. The novel amino acid sequences resulting from these changes are shown above each insert in upper-case letters (termination codons shown by an asterisk), with lower-case letters showing the point of transition from the amino acid sequences given in Figure 1. The arrow above SRp40-3 marks the position of the deletion of Arg185 in this clone, which may reflect the utilization of a different 3′ splice site. GenBank accession numbers: U30825–U30829, U30883 and U30884.

Constitutive splicing activity of recombinant SR proteins

The specific activities of the purified recombinant SR proteins were measured in a biochemical complementation assay based on the requirement for at least one SR protein for general splicing (Krainer *et al.*, 1990). A HeLa cell cytosolic S100 extract that contains all other required splicing components was complemented with individual recombinant SR proteins, and the splicing of a β-globin pre-mRNA was measured (Figure 5). In the case of SF2/ASF, the RS domain is essential for activity in this assay (Cáceres and Krainer, 1993; Zuo and Manley, 1993). Therefore, we were particularly interested in determining the activity of SRp30c, which has the shortest RS domain of all the SR proteins characterized to date. No splicing activity was detected in the absence of SR proteins (lane 1), whereas efficient splicing was obtained upon the addition of recombinant SRp30c, SRp40 and SRp55 (lanes 8–16), whose activities were comparable with those of SF2/ASF and of baculovirus-expressed SC35 (lanes 2–7). Slight differences in specific activities may be caused by different renaturation efficiencies or each SR protein may have preferred pre-mRNA substrates. Unfortunately, SRp20 and SRp75, like SC35, were not expressed in *E.coli* and could not be included in this assay. With some of the proteins, particularly SRp30c, the largest amounts inhibited splicing, resulting in the partial degradation of the pre-mRNA (lane 10). This may be a non-specific effect caused by, for example, residual bacterial RNA.

In vivo analysis of SR protein function in alternative splicing

In addition to their role in constitutive splicing, SR proteins can modulate the selection of alternative splice sites, as mentioned above. We assayed the three new human SR proteins for alternative splicing activity by a transient cotransfection assay (Cáceres *et al.*, 1994). Each of the SR protein cDNAs under the control of a cytomegalovirus (CMV) promoter was cotransfected with one of two reporter genes, and the different alternatively spliced reporter mRNAs were analysed by RT-PCR.

We first used a thalassemic allele of the human β-globin gene whose splicing is responsive to changes in the levels of SF2/ASF both *in vitro* and *in vivo* (Krainer *et al.*, 1990; Cáceres *et al.*, 1994). This thalassemic allele contains a G → A transition that disrupts the invariant GT dinucleotide at the 5′ splice site of intron 1. The mutation at this 5′ splice site activates three cryptic splice sites both *in vitro* and *in vivo* (Treisman *et al.*, 1983; Krainer *et al.*, 1984). The cotransfection of different SR cDNAs (Figure

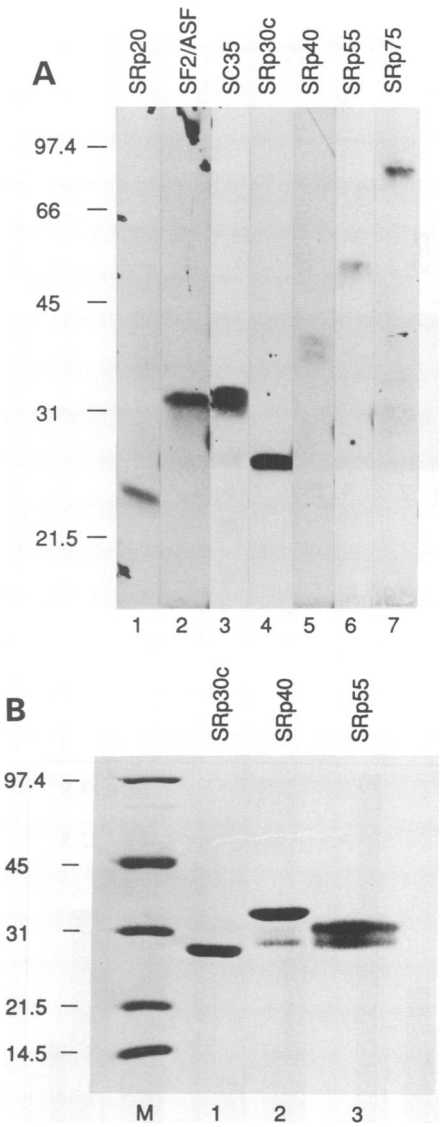


Fig. 4. Recombinant SR proteins. **(A)** [^{35}S]methionine-labelled SR proteins generated by coupled *in vitro* transcription and translation in rabbit reticulocyte lysate. The products were detected by autoradiography following SDS–PAGE. **(B)** Coomassie blue-stained SDS gel of purified recombinant SR proteins expressed in *E.coli*. The faster migration of these proteins, compared with (A), is probably a result of the lack of RS domain phosphorylation in bacteria.

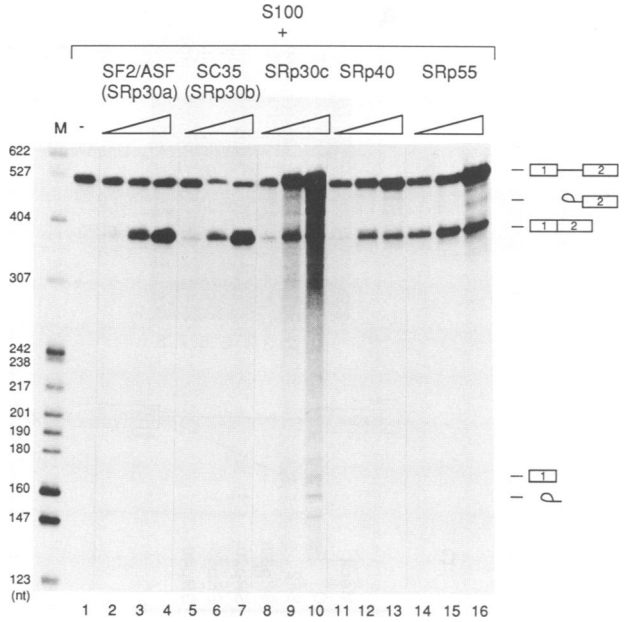


Fig. 5. *In vitro* analysis of constitutive splicing activity of recombinant SR proteins. β-globin pre-mRNA was incubated under splicing conditions in HeLa cell S100 extract alone (lane 1) or S100 extract complemented with increasing amounts (0.2, 0.4 and 0.8 μM) of the indicated SR proteins (lanes 2–16). The structures of the pre-mRNA and splicing products are shown schematically on the right, and the sizes of the pBR322/*Hpa*II markers (lane M) are shown on the left.

6A) resulted in changes in cryptic 5′ splice site use that were characteristic for each SR protein. Compared with pCG vector control (lane 4), SF2/ASF gave the strongest activation of the most proximal site, cryptic site 3 (lane 3), as described previously *in vivo* and *in vitro* (Krainer *et al.*, 1990; Cáceres *et al.*, 1994). SRp20 had a similar effect (lane 6), whereas SRp30c showed a slight activation of cryptic sites 1 and 3 (lane 5). Little to no splice site switching was observed with SRp40 and SRp55 in several experiments (Figure 6A, lanes 7 and 8, and data not shown).

To examine the possibility that the lack of activity of some of these proteins resulted from a failure to express the transfected cDNA constructs, we examined their expression by RT-PCR, using an upstream primer specific for the vector 5′ UTR and a downstream primer specific for each SR protein (Figure 6B). In each case, expression of the transfected cDNA was detected and we could show that the signal was not caused by the amplification of residual plasmid DNA because it was dependent on the inclusion of reverse transcriptase. Although we cannot determine the relative levels of each SR protein in its active form, it is clear that each of the cDNAs was expressed in the transfected cells.

A similar cotransfection experiment was carried out using an adenovirus E1A reporter construct (Figure 6C). RT-PCR was carried out with a labelled primer and a reduced number of cycles to allow the quantitation of the data by PhosphorImage analysis (Figure 6D). In previous *in vitro* studies, calf thymus SRp30b, SRp40, SRp55 and SRp70, as well as human SF2/ASF, were shown to stimulate the use of the 13S 5′ splice site (Harper and Manley, 1992; Mayeda and Krainer, 1992; Zahler *et al.*, 1993a). In our transient transfection study, SF2/ASF strongly activated the most proximal 5′ splice site, generating the 13S mRNA (Figure 6C, lane 3, compare with control in lane 2), as described previously (Cáceres *et al.*, 1994). A more modest shift towards 13S splicing was also obtained with SRp30c and SRp40 (compare the ratio of 13S:12S mRNA in lanes 4 and 7 with that in lane 2; see also Figure 6D), and less so with SRp55 (lane 6). Remarkably, SRp20 reproducibly inhibited the use of the 13S and 9S 5′ splice sites and stimulated the use of the 12S site (lane 5). Thus, SRp20 activated the proximal cryptic 5′ splice site of the β-thalassemia reporter and the middle 5′ splice site of the E1A reporter, whereas SF2/ASF activated the proximal 5′ splice site of both reporters. This observation demonstrates that different SR proteins can have substrate-specific effects on alternative splicing *in vivo*.

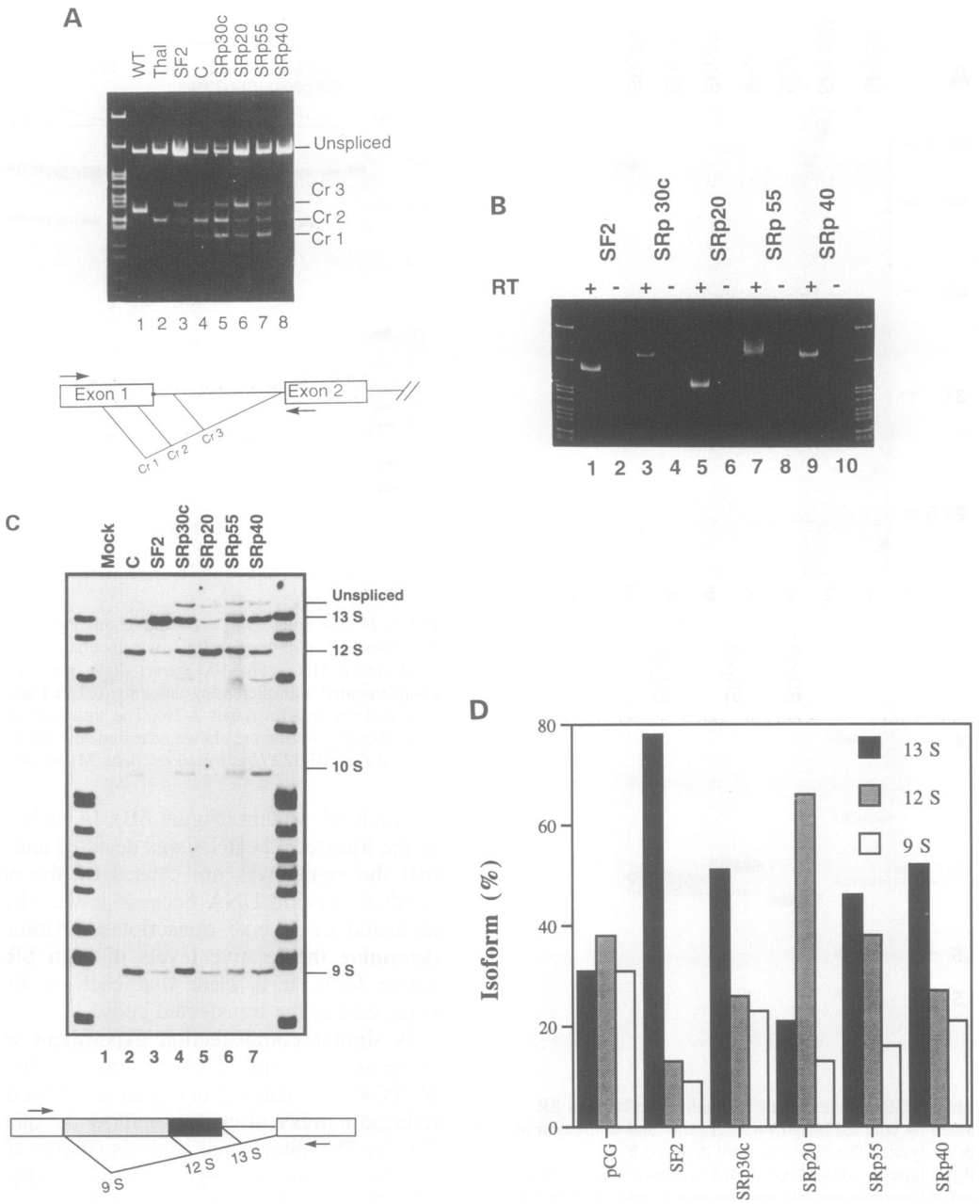


Fig. 6. *In vivo* analysis of alternative splicing activity of cloned SR proteins. HeLa cells were transiently cotransfected with a thalassemic β-globin gene or an adenovirus E1A gene and one of the indicated SR protein cDNAs subcloned in the pCG expression plasmid. Total RNA was analysed by RT-PCR followed by urea–PAGE. Detection was by using ethidium bromide staining (A and B) or a 5′ end-labelled primer and autoradiography (C). (**A**) Effect of SR protein overexpression on the alternative splicing of β-thal pre-mRNA. Lanes 1 and 2, control transfections with wild-type or thalassemic β-globin genes only, respectively; lane 4, control cotransfection with pCG lacking an insert; lanes 3 and 5–8, cotransfections with the indicated SR protein cDNAs. The mobilities of unspliced pre-mRNA and of the alternatively spliced mRNAs corresponding to the use of each of the three cryptic 5′ splice sites are indicated on the right. The diagram below shows the structure of the mutant allele and the positions of the cryptic 5′ splice sites, as well as the location of the PCR primers. (**B**) Expression of transfected SR protein cDNAs. Aliquots of total RNA from the cells cotransfected in (A) were analysed by RT-PCR, with one primer corresponding to each of the indicated SR proteins and a common primer corresponding to the vector 5′ UTR. Reverse transcriptase was omitted from the reactions in the even-numbered lanes. (**C**) Effect of SR protein overexpression on adenovirus E1A alternative splicing. pBR322/*Hpa*II DNA size markers are shown on both sides; lane 1, mock transfection; lane 2, control cotransfection with pCG lacking an insert; lanes 3–7, cotransfections with the indicated SR protein cDNAs. The mobilities of unspliced pre-mRNA and of 13S, 12S, 10S and 9S mRNAs are indicated on the right. The diagram below shows the structure of the E1A gene and of the major mRNAs generated by alternative 5′ splice site selection. The 10S and 11S mRNAs arise from double splicing events (Stephens and Harlow, 1987) and are not shown in the diagram. (**D**) Quantitation of the relative use of alternative E1A 5′ splice sites in response to SR protein overexpression. The data in (C) were quantitated by PhosphorImage analysis, and the amounts of 13S, 12S and 9S mRNAs are expressed as a percentage of the sum of all three mRNA isoforms.

Changes in expression of SR protein genes upon T cell activation

We wished to determine whether the expression of the newly identified SR proteins changes under conditions in which alternative splicing of many cellular genes undergoes regulated changes. The mitogenic stimulation of T cells constitutes a useful experimental system in which to study alternative splicing regulation in response

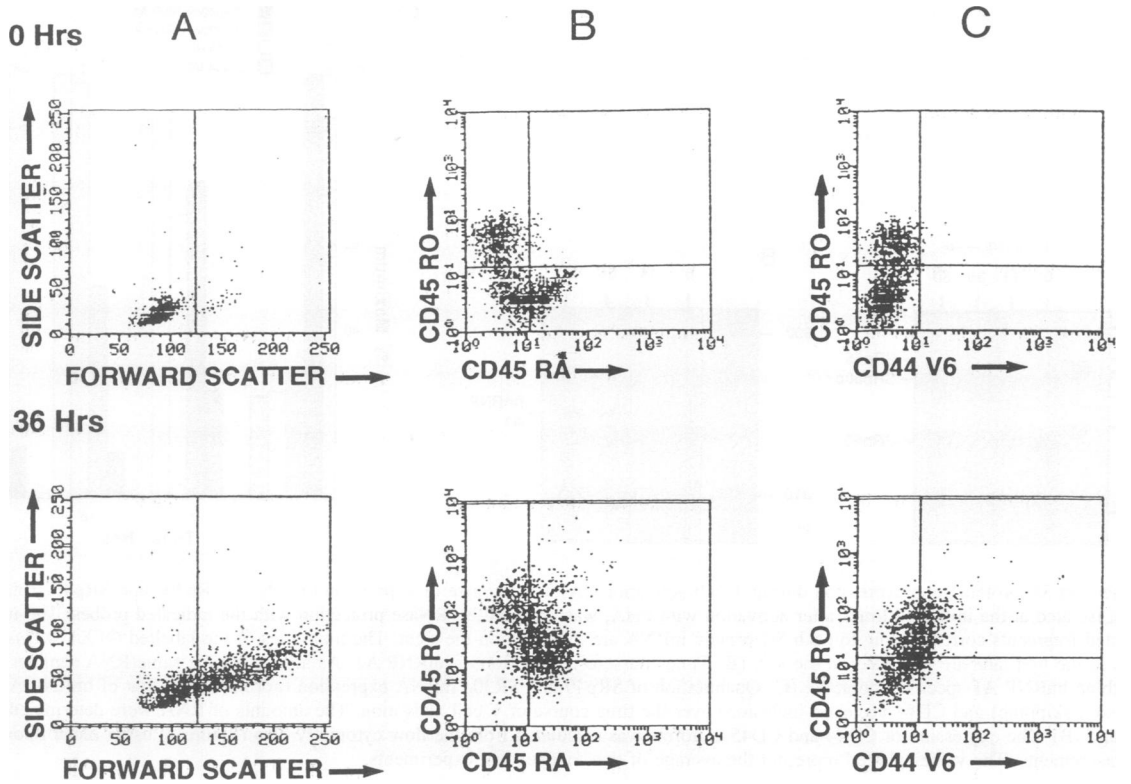


Fig. 7. Time course of T cell activation and expression of alternatively spliced variants of CD44 and CD45. T cells were analysed by flow cytometry at time 0 (top) and 36 h (bottom) following PHA stimulation. (**A**) Forward scatter to follow the appearance of larger T lymphoblasts. (**B**) Double staining with anti-CD45 RO (specific for the CD45 isoform in which exons 4–6 are skipped and exon 3 is joined to exon 7) and anti-CD45 RA (specific for CD45 isoforms that include exon 4). (**C**) Double staining with anti-CD45 RO and anti-CD44 v6 (specific for isoforms that include the optional exon v6).

to changes in cell physiology. Following T cell activation by phytohaemagglutinin (PHA), defined changes take place in several alternatively spliced pre-mRNAs, including the cell surface molecules CD44 and CD45, as the cells enter the cell cycle (Deans *et al.*, 1989; Jackson *et al.*, 1994). A prerequisite for the involvement of one or more SR proteins in this switch is that the levels, post-translational modifications or accessibility of the SR protein, or of one of its antagonists, should vary during the same time course.

PHA-treated T cells were analysed by flow cytometry using antibodies specific for the alternatively spliced isoforms of CD44 and CD45 (Figure 7). Anti-CD44 v6 recognizes CD44 isoforms that include the optional exon v6 (Fox *et al.*, 1994). Two antibodies were used to detect $CD45^+$ cells: (i) anti-CD45 RO recognizes an epitope comprising residues encoded in exons 3 and 7 when these residues are adjacent, as a result of the skipping of exons 4–6 (Schmidt, 1989); and (ii) anti-CD45 RA recognizes isoforms that include the optional exon 4. At 36 h post-stimulation, larger T cell blasts appear (Figure 7A), concomitant with the increased expression of both CD45 RO (Figure 7B; exon skipping) and CD44 v6 (Figure 7C; exon inclusion) isoforms. Figure 7C also shows that the CD45 RO and CD44 v6 isoforms are coexpressed on a number of cells at 36 h post-stimulation (upper right quadrant), demonstrating that two different pre-mRNAs can simultaneously undergo opposite patterns of exon inclusion or skipping in individual cells.

To measure the expression of the genes encoding the three newly identified human SR proteins, total RNA isolated at various times following T cell activation was analysed by RNase protection with a mixture of SRp30c, SRp40 and SRp55 cDNA riboprobes (Figure 8A). The expression of hnRNP A1, a known antagonist of several, if not all, SR proteins (Mayeda and Krainer, 1992; Mayeda *et al.*, 1994), was also measured (Figure 8B). Following PHA addition, marked changes were observed in the relative abundance of the different SR protein mRNAs (Figure 8A), while the levels of hnRNP A1 mRNA remained relatively constant (<25% variation; Figure 8B). SRp40 mRNA decreased markedly, whereas SRp55 decreased more gradually and SRp30c increased later on. The probe used for SRp55 bridged the alternatively spliced region at the 3′ end of the SRp55-3 cDNA clone. No full-length protection (347 bp) was observed for this probe, which was reduced to the 275 bp fragment common to all SRp55 isoforms, indicating that the SRp55-3 isoform is not expressed in T cells. The above changes in SR protein gene expression coincided with changes in CD44 and CD45 alternative splicing patterns (Figure 8C). Further experiments will be necessary to determine if these and/or other SR proteins are indeed responsible for the changes in CD44 and CD45 alternative splicing during T cell activation.

Discussion

Conserved structural features of SR proteins

We have isolated several human cDNAs representing three new members of the SR protein family of pre-mRNA

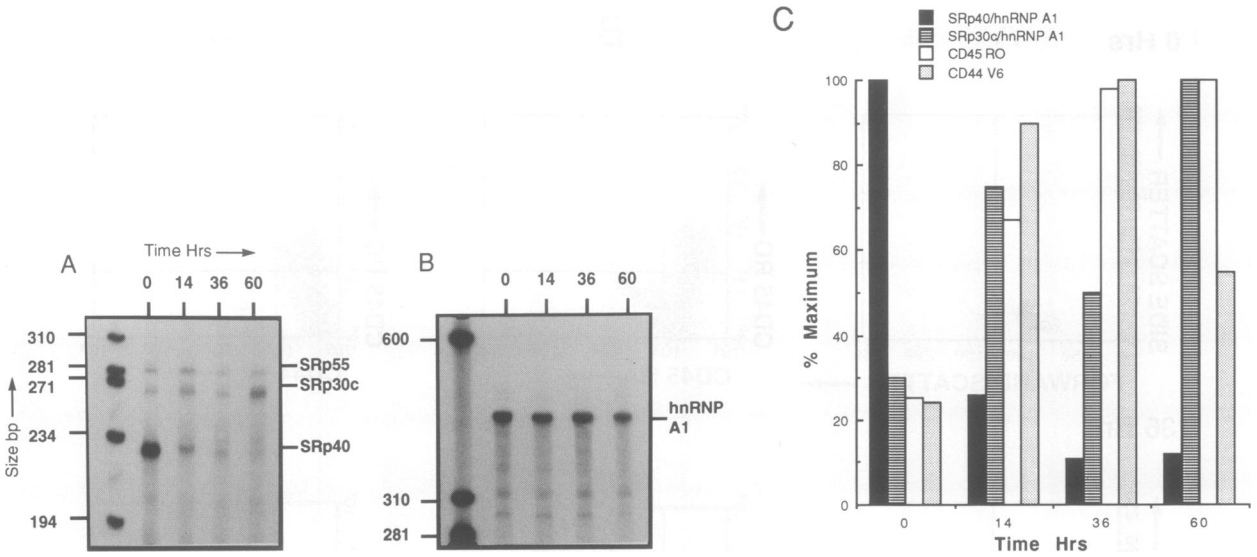


Fig. 8. Changes in SR protein gene expression during T cell activation. **(A)** Time course of expression of SRp55, SRp30c and SRp40. Total RNA from T cells, isolated at the indicated times after activation with PHA, was analysed by RNase protection with the indicated probes. The mobilities of the protected fragments corresponding to each SR protein mRNA are indicated on the right. The mobilities of the labelled ΦFX174/*Hae*III DNA size markers in the first lane are indicated on the left. **(B)** Time course of expression of hnRNP A1. As above, but the same RNA samples were analysed with an hnRNP A1-specific riboprobe. **(C)** Quantitation of SRp40 and SR30c mRNA expression (normalized to that of hnRNP A1) and of CD45 RO (exon skipping) and CD44 v6 (exon inclusion) over the time course of T cell activation. The amounts of RNA were determined from the data in (A) and (B). The expression of CD44 and CD45 isoforms was calculated from the flow cytometry data (Figure 7) using mean fluorescence intensity measurements. The values plotted represent the average of two independent experiments.

splicing factors and alternatively spliced isoforms thereof. This raises the total number of known human SR proteins to eight, not including the alternatively spliced isoforms (Figure 9). Homologues and/or additional family members have been identified in numerous other eukaryotes (reviewed in Birney *et al.*, 1993). Human SRp40 is the family member of rat SRp40/HRS, whereas SRp55 is the homologue, of the known human SR proteins, most closely related to *Drosophila* SRp55/B52. The two proteins are very similar in size and are 64% identical at the amino acid level. Human SRp30c is a new member of the SR family. It resembles SF2/ASF most closely (74% amino acid identity) and has an unusually short RS domain.

SR proteins fall into two structural classes, depending upon the presence of a central atypical RRM. Of the human SR proteins, SC35, SRp20 and 9G8 lack the central RRM. The three family members we report contain the central RRM with its invariant signature SWQDLKD (Figure 9A), as expected from the PCR screen used to isolate their cDNAs. Few other proteins or open reading frames in the sequence databases contain this signature, and in all cases it is located at a homologous position within an RRM (Birney *et al.*, 1993). All of these proteins have an additional N-terminal RRM, and the overall homology of the two tandem RRMs to bona fide SR proteins is much higher than expected for randomly selected RRMs. However, not all known proteins with the above signature have a clear C-terminal RS domain with consecutive Arg–Ser dipeptides (e.g. NPL3 in *Saccharomyces cerevisiae*; Bossie *et al.*, 1992; Birney *et al.*, 1993), which is the other characteristic feature of all SR proteins.

The close sequence relatedness of the different SR proteins can be seen in the alignment of their RRMs (Figure 9A). As for the other SR proteins with two RRMs, the central RRM of each of the newly identified human SR proteins has only a limited match to the consensus hexamer RNP-2 and octamer RNP-1 submotifs, whereas the N-terminal RRMs contain clear matches to these sequences. All of the central RRMs have a closer sequence resemblance to each other than to the N-terminal RRMs (data not shown), suggesting that domain duplication to generate this central RRM preceded the gene duplications that gave rise to different SR family members (Birney *et al.*, 1993). Both types of RRM are thought to have a similar tertiary structure (Kenan *et al.*, 1991; Birney *et al.*, 1993). The size differences among SR proteins arise from the presence or absence of the central RRM and from the variable length of the C-terminal RS domain (Zahler *et al.*, 1992; Birney *et al.*, 1993; Figure 9B).

Alternatively spliced isoforms of SR proteins

We have isolated a number of cDNAs encoding variants of SRp40 and SRp55, which most likely result from several forms of alternative splicing, including exon skipping, intron retention and the use of alternative 5′ or 3′ splice sites. Other alternatively spliced isoforms of SR proteins have been reported: *Drosophila* SRp55 (Roth *et al.*, 1991) and B52 (Champlin *et al.*, 1991) appear to be closely related variants encoded by a single gene, which differ primarily by the presence or absence of a short alternative exon. In addition, alternatively spliced isoforms of human SF2/ASF (ASF-3) and *Drosophila* RBP1 lack the RS domain as a result of retention of an intron, while another SF2/ASF isoform, ASF-2, results from the use of an alternative 3′ splice site (Ge *et al.*, 1991; Kim *et al.*, 1992). Alternatively spliced SC35 isoforms have also been reported which differ in the 3′ UTR, leading to differences in the stability of the various

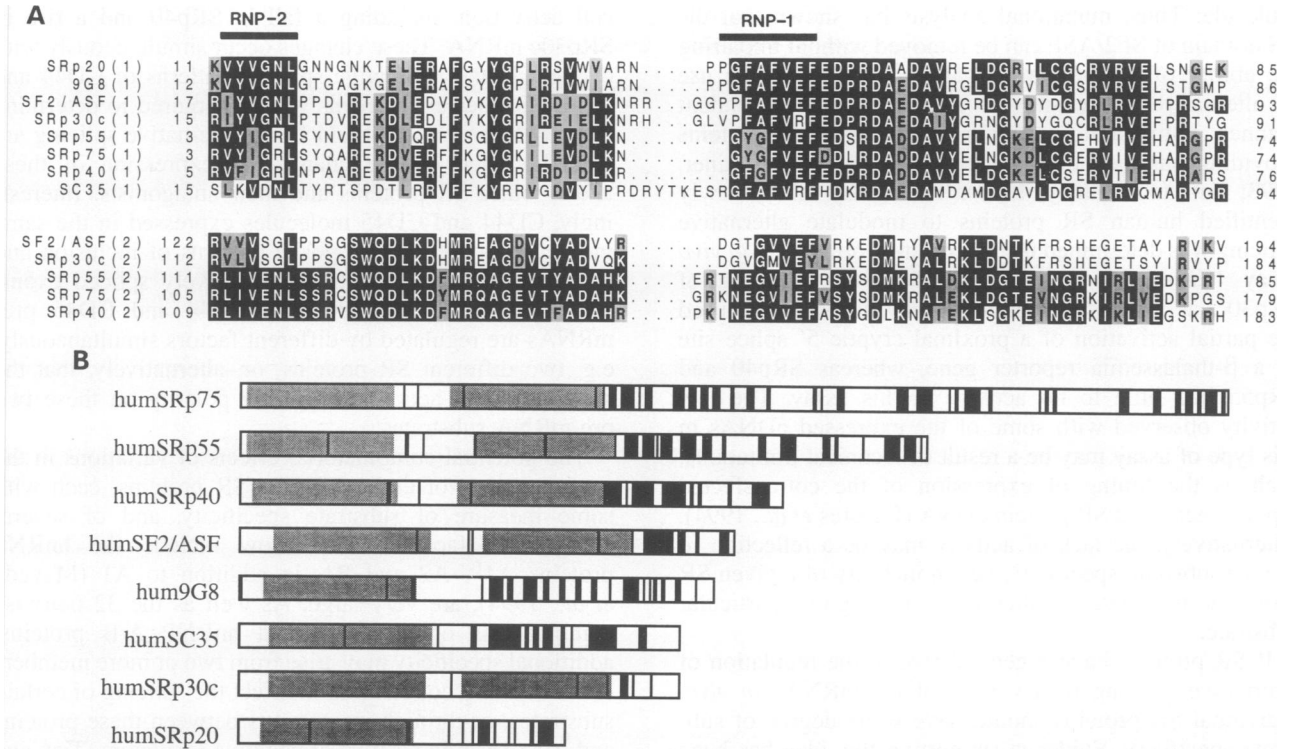


Fig. 9. Sequence homology and domain structure of human SR proteins. (**A**) Alignment of RRM1 and RRM2 sequences encoded by the newly isolated cDNAs, along with RRMs of other human SR proteins. The positions of the conserved RNP-2 and RNP-1 submotifs are indicated above the alignment. The RRM2 sequences have poor matches to these submotifs, but are thought to have the same overall tertiary fold (Birney *et al.*, 1993). Residue conservation was computed separately for RRM1 (top) and RRM2 (bottom), although several residues are conserved at equivalent positions in both alignments. The positions of conserved residues are highlighted by vertical shading. Black shading indicates positions in which a single residue occurs in more than half of the sequences. Grey shading at the same positions represents conservative substitutions. Elsewhere, grey shading indicates positions in which residues belonging to a single conservative grouping (I = V = L, F = Y, R = K, D = E, G = A) are present in more than half of the sequences. (**B**) Schematic representation of domain structures of all eight human SR proteins drawn to scale. Grey shading represents RRMs. Thin dark vertical lines represent individual RS or SR dipeptides. Thicker lines indicate clustering of RS or SR dipeptide repeats, with the thickness being proportional to the number of consecutive repeats (Birney *et al.*, 1993). The human SR protein sequences are taken from the references listed in the text, or from Figure 1.

transcripts (Sureau and Perbal, 1994). Significantly, the expression of alternatively spliced SR protein isoforms is phylogenetically conserved, as the same three isoforms of SF2/ASF have been reported in mouse (Tacke *et al.*, 1992) and two of the alternatively spliced SRp40 isoforms we have reported here were also found in HRS, the rat homologue of SRp40 (Diamond *et al.*, 1993). Alternative splicing is also used as a mechanism for regulating the expression of other splicing factors, such as SWAP, tra-2 and hnRNP A1/A1^B and A2/B1 (reviewed in Mattox *et al.*, 1992).

Alternative splicing as seen in the SRp40-2, SRp40-3 and SRp55-2 cDNA clones would lead to truncated proteins containing RRM1 alone. In the case of SF2/ASF, whose primary structure is very similar to that of SRp40 and SRp55, RRM1 alone does not bind RNA and is inactive in splicing (Cáceres and Krainer, 1993). The longer recombinant proteins expressed from the ASF-2 and ASF-3 cDNAs actually inhibit splicing (Ge *et al.*, 1991). At present, it is unclear if the alternatively spliced mRNAs encoding variant SR proteins that lack one or more of the characteristic domains arise from inefficient splicing (in the case of intron retention), reflect a post-transcriptional control mechanism for reducing SR protein production and/or encode stable protein isoforms with distinct functions.

Splicing activities of human SRp30c, SRp40 and SRp55

Recombinant SRp30c, SRp40 and SRp55 were all produced in *E.coli*. Individually, these recombinant proteins are able to restore splicing to a cytosolic S100 extract deficient in all SR proteins. Previous studies of SF2/ASF demonstrated that the RS domain is required for this constitutive splicing activity, that both Arg and Ser residues are required and that eight consecutive RS dipeptides are sufficient (Cáceres and Krainer, 1993; Zuo and Manley, 1993). SRp30c, which among the SR proteins is the most closely related to SF2/ASF in primary structure (74% amino acid identity), was active in the complementation assay, although its specific activity was consistently lower than that of SF2/ASF. This lower specific activity may be a result of the size of the SRp30c RS domain, which is the shortest of all the SR proteins reported to date, with 7 Arg residues, 10 Ser residues and only 5 SR dipeptides. In contrast, the RS domain of wild-type SF2/ASF has 13 RS or SR dipeptides (Ge *et al.*, 1991; Krainer *et al.*, 1991). It is possible that SRp30c functions optimally in constitutive splicing only with certain pre-mRNA substrates.

The role of SR proteins in alternative splicing appears to be distinct from their constitutive splicing function and may involve different, albeit overlapping, domains of the

molecule. Thus, mutational analysis has shown that the RS domain of SF2/ASF can be removed without impairing the ability of the protein to cause switching in the use of alternative splice sites in a concentration-dependent manner, with limiting amounts of other SR proteins providing basal splicing function (Cáceres and Krainer, 1993; Zuo and Manley, 1993). The ability of the newly identified human SR proteins to modulate alternative splicing was determined using a recently developed *in vivo* assay (Cáceres *et al.*, 1994). The overexpression of SRp30c, as well as that of SF2/ASF and SRp20, caused the partial activation of a proximal cryptic 5′ splice site in a β-thalassemia reporter gene, whereas SRp40 and SRp55 had little to no activity in this assay. The low activity observed with some of the expressed cDNAs in this type of assay may be a result of technical limitations, such as the timing of expression of the cotransfected reporter gene and SR protein cDNA (Cáceres *et al.*, 1994). Alternatively, the lack of activity may be a reflection of *in vivo* substrate specificity, i.e. an inability of a given SR protein to modulate the alternative splicing of a particular substrate.

If SR proteins have a general role in the regulation of alternative splicing of a variety of pre-mRNAs *in vivo*, individual SR proteins should have some degree of substrate specificity. Evidence supporting this idea has been obtained with several substrates *in vitro* (Kim *et al.*, 1992; Fu, 1993; Sun *et al.*, 1993; Tian and Maniatis, 1993; Zahler *et al.*, 1993a; Staknis and Reed, 1994). The experiments reported here revealed differential substrate-specific effects of individual SR proteins *in vivo*. This differential specificity was particularly apparent with the adenovirus E1A reporter, which gave rise predominantly to 13S mRNA when cotransfected with SF2/ASF (and to a lesser extent with SRp30c, SRp40 and SRp55), whereas the 12S mRNA became the predominant species when SRp20 was overexpressed. In addition, SRp20 activated the proximal 5′ splice site of one reporter and the middle 5′ splice site of a different reporter, a further indication of specificity. We also note that this is the first demonstration of splicing activity for mammalian SRp20.

Substrate specificity most probably arises from differences in sequence-specific RNA binding (Sun *et al.*, 1993), although the determinants of binding specificity remain to be defined. The precise function of the second RRM, which is required in some SR proteins and is absent in others, and the significance of the wide variation in the length of their RS domains, also remain unknown. Although a short RS domain appears to be sufficient for splicing function in the case of SRp30c or of mutant SF2/ASF, the lengths and sequences of the RS domains are highly conserved between homologous family members in different species, arguing for highly specific roles of each RS domain with at least some natural substrates *in vivo*.

A prerequisite for a broad role of SR proteins in tissue-specific and developmentally regulated alternative splicing is that the expression or activity of individual family members or their antagonists must vary between tissues or developmental stages in which relevant target genes display varying patterns of alternative splicing. We have shown that complex changes in the levels of mRNA encoding the newly identified SR proteins occur upon T cell activation, including a fall in SRp40 and a rise in SRp30c mRNA. These changes occur simultaneously with changes in the alternative splicing patterns of CD44 and CD45. Further experiments will be required to determine if the changes in CD44 and CD45 alternative splicing are indeed caused by the changes in expression of these and/or other SR proteins and their antagonists. Interestingly, CD44 and CD45 molecules expressed in the same individual cell exhibit opposite patterns of inclusion and skipping of their respective alternatively spliced exons. This observation suggests that CD44 and CD45 pre-mRNAs are regulated by different factors simultaneously, e.g. two different SR proteins, or, alternatively, that the same regulator acts with opposite polarity on these two pre-mRNA substrates.

The potential combinatorial effects of variations in the concentrations of eight different SR proteins, each with some measure of substrate specificity, and of several antagonistic factors, which now include the hnRNP proteins $A1^B$, A2 and B1, in addition to A1 (Mayeda *et al.*, 1994), are very large. As well as the 32 pairwise combinations of these SR and hnRNP A/B proteins, additional specificity may arise from two or more members of each family cooperating to affect the splicing of certain substrates, or through interactions between these proteins and gene-specific positive or negative regulators (Tian and Maniatis, 1993). When yet more variables, such as the regulation of SR protein accessibility or localization (Jiménez-García and Spector, 1993; Cavaloc *et al.*, 1994; Gui *et al.*, 1994) and the potential regulation of SR protein activity by changes in phosphorylation (Cardinali *et al.*, 1994; Gui *et al.*, 1994), are added into the equation, the system gains a complexity perhaps sophisticated enough to control the numerous alternative splicing choices made in living cells.

Materials and methods

Cloning of cDNAs encoding new human SR proteins

Four degenerate primers encoding conserved sequence motifs were synthesized, with inosine in the positions of four-base degeneracy and with *Eco*RI or *Xba*I sites at the 5′ ends. The oligonucleotide sequences were as follows: RRM1a-*Eco* (RDADDA), TGCGAATTCCGIGAYGCIGAYGAYGC; RRM1b-*Eco* (DADDAVYE) TGAGAATTCGAYGCIGAYGAYGCIGTITAYGA; RRM2a-*Xba* (MRQAGEV), CTATCTAGACYTCICCIGCYTGICGCAT; and RRM2b-*Xba* (WQDLKDY/HM), CCATCTAGACATRTRRTCYTTIARRTCYTGCCA. Four PCRs with the primer combinations RRM1a-*Eco*/RRM2a-*Xba*, RRM1a-*Eco*/RRM2b-*Xba*, RRM1b-*Eco*/RRM2a-*Xba* and RRM1b-*Eco*/RRM2b-*Xba* were carried out with a mixed pool of cDNA templates derived from three human cell lines, HT29, HeLa and MOLT-4. For each cell line, first-strand cDNA was prepared by oligo(dT) priming of total RNA. PCR amplifications were performed in a 100 μl reaction mix that contained 10 mM Tris–HCl, pH 8.4, 50 mM KCl, 2.5 mM $MgCl_2$, 250 μM dNTPs, the appropriate primers at 400 nM and 4 U *Taq* DNA polymerase. The cycle parameters were 95°C for 5 min, 50°C for 1 min and 72°C for 90 s (one cycle), followed by 95°C for 1 min, 50°C for 1 min and 72°C for 90 s for 40 cycles. All four reactions generated products of the expected sizes, which were cloned into M13-mp18 following digestion with *Eco*RI and *Xba*I. Colony lifts on Hybond N^+ (Amersham, UK) were probed with 5′ ^{32}P end-labelled oligonucleotides specific for SRp75 and SF2/ASF to exclude these clones. Single-stranded templates were sequenced from 48 colonies.

Two oligonucleotides were made for each of three new sequences identified. These oligonucleotides were used to probe duplicate lifts of 2×10^5 colonies of a human HT29 (colonic carcinoma) cDNA library in pCDM8 (a gift from D.Simmons) electroporated into the strain MC1061-p3. Colonies that gave positive signals with both oligonucleotides were

isolated, the plasmids were restriction mapped and the inserts were subcloned into the *Xba*I site of M13-mp18. Single-stranded templates with the inserts in both orientations were sequenced using the chain termination technique with Sequenase (USB) and internal sequencing primers.

DNA sequence analysis

DNA analysis was performed using the GCG programs. Sequence comparisons used the pairwise alignment method (Needleman and Wunsch, 1970) with the gap program to deduce similarity and identity scores and the pile-up program to create the alignment in Figure 9A. The most similar sequences are grouped closer together in the alignment. The boxed alignment output and the domain structure diagram (Figure 9B) were generated using the PostScript programs kindly provided by J.Posfai and E.Birney, respectively.

Northern blot analysis

A Hybond N^+ filter containing 40 μg total RNA from selected mouse tissues (a kind gift from D.Blake) was probed with a random-primed PCR fragment of SRp30c amplified with the primers FSRp30c-*Xba* and RSRP30c-*Bam* (see below). After autoradiography, the filter was stripped and reprobed with a similarly generated mouse β-actin probe.

Expression and purification of recombinant SR proteins

cDNA clones encoding the three new SR proteins were amplified using the following primers: FSRp30c-*Nde*, CGGGCGGTGCATATGTCGGGCTGG; RSRp30c-*Bam*, GAGGGATCCTCAGTAGGGCCTGAAAGGAGAG; FSRp40-*Nde*, CTAGCCGGACCATATGAGTGGCTGTC; RSRp40-*Bcl*, GGCTGGTGATCATTAATTGCCACTGTCAACTGATCTGG; FSRp55-*Nde*, CCCGCCACGCATATGCCGCGCGTC; and RSRp55-*Bam*, GACGGATCCTTAATCTCTGGAACTCGACCTGGAC. Following a 20 cycle amplification using Pfu polymerase (Stratagene), the products were digested and cloned into the *Nde*I and *Bam*HI sites of the T7 RNA polymerase bacterial expression vector pET-9c (Novagen). The sequences of these clones were checked by single-stranded sequencing of inserts transferred into M13-mp18. Integrity of the clones was also checked in a coupled transcription–translation assay using T7 RNA polymerase and rabbit reticulocyte lysate, according to the manufacturer's protocol (Promega).

Following transformation of the expression plasmids into *E.coli* strain BL21(DE3)pLysS (Novagen), single colonies were grown in 1 l of M9ZB medium (Studier *et al.*, 1990) supplemented with 25 mg/l kanamycin and chloramphenicol. Cells were induced with 0.4 mM IPTG at a density of 0.4 $OD_{550\ nm}$ and grown for a further 3 h.

The bacterial cells were lysed by sonication in buffer D (100 mM KCl, 20 mM HEPES/Na^+, pH 8.0, 0.2 mM EDTA, 5% glycerol) supplemented with 1 mM dithiothreitol and 0.5 mM phenylmethylsulfonyl fluoride. Then they were centrifuged at 30 000 *g* for 15 min. For the purification of SRp40, the inclusion bodies were first washed in 1% sodium deoxycholate, then in buffer D, and finally dissolved in buffer D containing 6 M urea. SRp30c was partially soluble and was recovered from the soluble fraction by precipitation in the presence of 20 mM $MgCl_2$. The pellet remained insoluble in buffer D containing 20 mM EDTA, so instead it was dissolved in buffer D without KCl and with 6 M urea. SRp55 was expressed at lower levels than SRp30c and SRp40, and hence purification started with a 20 l culture. Following lysis by sonication, the soluble fraction was loaded onto a CsCl density gradient to remove nucleic acids, as described previously (Krainer *et al.*, 1991). Protein-containing fractions were dialysed in buffer D without salt, resulting in the selective precipitation of SRp55, which was then dissolved in buffer D with 6 M urea. The same method was used to prepare recombinant SF2/ASF (Krainer *et al.*, 1991).

To remove minor contaminants, each recombinant SR protein was purified further by Mono S chromatography (Pharmacia) in buffer D with 6 M urea, and elution with a 0.1–1.0 M KCl gradient in the same buffer. SRp30c, SF2/ASF, SRp40 and SRp55 eluted at ~150, 400, 450 and 550 mM KCl, respectively. The fractions containing SRp40 were pooled, dialysed into buffer D with 6 M urea, and purified further by chromatography on Mono Q using the same buffers as with Mono S. SRp40 eluted at ~450 mM KCl. Each of the proteins was then dialysed into buffer D, except for SRp30c, which became insoluble and hence was stored in buffer D with 6 M urea.

Purified recombinant SC35 expressed from a baculovirus vector in Sf9 cells (Fu and Maniatis, 1992b) was a generous gift from K.Lynch and T.Maniatis.

In vitro splicing

In vitro splicing assays were carried out in 25 μl reactions with 20 fmol of ^{32}P-labelled, $7CH_3$-GpppG-capped SP6 transcripts of β-globin pre-mRNA generated from the plasmid SP64-HβΔ6 linearized at the *Bam*HI site in exon 2 (Krainer *et al.*, 1984; Mayeda and Krainer, 1992). The reactions contained 8 μl of HeLa cell S100 extract in the above buffer D (Krainer *et al.*, 1990) complemented with 0, 5, 10 or 20 pmol of recombinant SR protein in buffer D. The stock of SRp30c in buffer D with 6 M urea was diluted in buffer D immediately before its addition to the splicing reaction. To allow a comparison of the effects of different SR proteins at different concentrations, all the reactions were adjusted to a final concentration of 61 mM urea. Following incubation for 5 h at 30°C, the RNA was extracted and analysed by urea–PAGE and autoradiography.

In vivo analysis of alternative splicing

SRp20, SRp30c, SRp40 and SRp55 cDNAs isolated from the HT29 cDNA library were amplified by PCR, as described above, with the following primer pairs: FSRp20-*Spe* (CGCACTAGTCATCGTGATTCCTGTCCGTTGGAT) and RSRp20-*Bam* (CGGGGATCCTATTTCCTTTCATTTGACCTAGATCG); FSRp30c-*Xba* (CGCTCTAGATCGGGCTGGGCGGACGAGCGC) and RSRp30c-*Bam* (see above); FSRp40-*Xba* (CGCTCTAGAAGTGGCTGTCGGGTATTCATC) and RSp40-*Bcl* (see above); and FSRp55-*Spe* (CGCACTAGTCCGCGCGTCTACATAGGACGC) and RSRp55-*Bam* (see above). The amplified DNAs were digested with the appropriate enzymes and subcloned into the *Xba*I and *Bam*HI sites of the mammalian expression vector pCG (Tanaka and Herr, 1990), under the control of the CMV enhancer/promoter. This procedure resulted in the insertion of the dipeptide Ser–Arg between the first two amino acids of each wild-type SR protein. For each SR protein cDNA, 1 μg of plasmid DNA was cotransfected into HeLa cells with 5 μg of the β-thalassemia reporter plasmid (Cáceres *et al.*, 1994), or with 6 μg of the adenovirus E1A expression plasmid pMTE1A (Zerler *et al.*, 1986) in the presence of 20 μg of lipofectin (Gibco-BRL). The cells were grown to 60–75% confluence in 60 mm dishes, harvested 48 h after transfection and 200 ng total RNA were analysed by RT-PCR, as described previously (Cáceres *et al.*, 1994). For the β-globin experiments and to measure the expression of transfected SR protein cDNAs, reverse transcription and 30 cycles of amplification were carried out with rTth polymerase; detection was by ethidium bromide staining. For the experiments with E1A, first-strand oligo(dT)-primed cDNA synthesized with AMV reverse transcriptase (Northumbria Biologicals) from 100 ng of total RNA was amplified for 25 cycles, using 5′ end-labelled forward primer. Detection and quantitation were carried out by autoradiography and PhosphorImage analysis (Fujix, BAS2000), respectively.

T cell fractionation and activation

Lymphocytes isolated from the buffy coat of 0.5 l of human blood were separated by density centrifugation (Lymphoprep, ICN Pharmaceuticals). T cells were purified from this preparation by passage over glass beads coated in goat anti-human IgG (Cellect, Biotex Inc.). A >95% pure population, as verified by fluorescence-activated cell sorting (FACS) analysis with anti-CD3 mAb, was stimulated with 50 μg/ml PHA (Wellcome) and 0.3% unfractionated cells to provide the necessary costimulation. Four aliquots of 5×10^7 cells at a density of 2×10^6 cells/ml were incubated at 37°C in RPMI 1640 medium containing 10% fetal calf serum for 0, 14, 36 and 60 h. At these time points, aliquots were taken for FACS analysis (Figure 7). The remaining cells from each sample were harvested and total RNA was extracted with RNAzol B (Cinna Biotecx Laboratories). The RNA was resuspended in hybridization buffer with 80% formamide (Melton *et al.*, 1984), and the concentration was determined based on the absorbance at 260 nm.

Flow cytometry

FACS analysis was carried out on a FACScan instrument (Becton Dickinson) with 2×10^5 T cells. The following antibodies were used: FITC-conjugated anti-CD45 RA (Immunotech, Marseille, France), phycoerythrin (PE)-conjugated anti-CD45 RO antibody, negative control PE- or FITC-conjugated antibodies, rabbit $F(ab)_2$ anti-mouse IgG FITC (all from Dakopatts, Denmark) and anti-CD44/v6 (British Biotechnology, Oxford, UK). Samples were blocked with 10% normal human serum or mouse serum for indirect detection. Most dead cells were eliminated by selecting the appropriate FSC/SSC gate. Compensation levels were determined on individual single-stained samples for all markers prior to the collection of double-stained samples. Markers for analysis were set

according to background readings, and were chosen so that no change would be required between time points.

RNase protection assays

Short segments of cloned cDNA encoding each of the SR proteins and hnRNP A1 were amplified by 20 cycles of PCR. A 268 bp fragment of SRp30c was generated with primers FSRp30c (CTGGATGACACC-AAATTCCGCTCT) and RSRp30c (CCTAAGACACTAAATCCTC-AATCTGG). For SRp40, primers FSRp40 (AAGTTCCTCTAGGTCT-CGTAGCCG) and RSRp40 (TTGCCACTGTCAACTGATCTGGAC) gave a 230 bp fragment. For SRp55, primers FSRp55 (AAGTCGC-TCCCGTTCCAGGTCG) and RSRp55-3 (AAGGGCAGCTTGATGCC-AGACTG), which hybridize to the alternatively spliced region in clone SRp55-3, gave a 347 bp fragment; 72 bp of this probe are specific to the alternatively spliced region of SRp55-3, and 275 bp are shared by all SRp55 isoforms. For hnRNP A1, primers FhnRNPA1 (GACACTG-AAGAACATCACCTAAGAG) and RhnRNPA1 (TGCCAAATCCAT-TATAGCCATCCC) generated a 383 bp fragment. The fragments were subcloned into the PCR II TA cloning vector (Invitrogen). [^{32}P]UTP-labelled antisense RNA probes were generated from *Bam*HI-linearized plasmids by transcription with SP6 RNA polymerase. 3×10^5 c.p.m. of each labelled probe were incubated overnight with 5 μg of total T cell RNA (Melton *et al.*, 1984). Following electrophoresis and autoradiography, individual bands were excised and the radioactivity was measured by scintillation counting.

Acknowledgements

We thank David Simmons for the gift of the HT29 cDNA library, Kristen Lynch and Tom Maniatis for purified SC35 and members of Peter Ratcliffe's laboratory for advice with the RNase protection assays. We are grateful to David Horowitz for comments on the manuscript. G.R.S. was an MRC Training Fellow and A.R.K. is a Pew Scholar in Biomedical Sciences. We would like to acknowledge the Wellcome Trust, the Arthritis and Rheumatism Council and the Medical Research Council for their support of G.R.S., M.V.B., M.P., D.G.J. and J.I.B., and the National Institutes of Health and the National Cancer Institute for their support of A.M., J.F.C. and A.R.K.

References

Ayane,M., Preuss,U., Kohler,G. and Nielsen,P.J. (1991) A differentially expressed murine RNA encoding a protein with similarities to two types of nucleic acid binding motifs. *Nucleic Acids Res.*, **19**, 1273–1278.

Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) RNA-binding proteins as developmental regulators. *Genes Dev.*, **25**, 431–437.

Birney,E., Kumar,S. and Krainer,A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.

Bossie,M.A., DeHoratius,C., Barcelo,G. and Silver,P. (1992) A mutant nuclear protein with similarity to RNA binding proteins interferes with nuclear import in yeast. *Mol. Biol. Cell.*, **3**, 875–893.

Cáceres,J.F. and Krainer,A.R. (1993) Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J.*, **12**, 4715–4726.

Cáceres,J.F., Stamm,S., Helfman,D.M. and Krainer,A.R. (1994) Regulation of alternative splicing by overexpression of antagonistic splicing factors. *Science*, **265**, 1706–1709.

Cardinali,B., Cohen,P.T.W. and Lamond,A.I. (1994) Protein phosphatase 1 can modulate alternative 5′ splice site selection in a HeLa splicing extract. *FEBS Lett.*, **352**, 276–280.

Cavaloc,Y., Popielarz,M., Fuchs,J.P., Gattoni,R. and Stévenin,J. (1994) Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. *EMBO J.*, **13**, 2639–2649.

Champlin,D.T., Frasch,M., Saumweber,H. and Lis,J.T. (1991) Characterization of a *Drosophila* protein associated with boundaries of transcriptionally active chromatin. *Genes Dev.*, **5**, 1611–1621.

Deans,J.P., Boyd,A.W. and Pilarski,L.M. (1989) Transitions from high to low molecular weight isoforms of CD45 (T200) involve rapid activation of alternate mRNA splicing and slow turnover of surface CD45R. *J. Immunol.*, **143**, 1233–1238.

Diamond,R.H., Du,K., Lee,V.M., Mohn,K.L., Haber,B.A., Tewari,D.S. and Taub,R. (1993) Novel delayed-early and highly insulin-induced growth response genes. Identification of HRS, a potential regulator of alternative pre-mRNA splicing. *J. Biol. Chem.*, **268**, 15185–15192.

Eperon,I.C., Ireland,D.C., Smith,R.A., Mayeda,A. and Krainer,A.R. (1993) Pathways for selection of 5′ splice sites by U1 snRNPs and SF2/ASF. *EMBO J.*, **12**, 3607–3617.

Fox,S.B., Fawcett,J., Jackson,D.G., Collins,I., Gatter,K.C., Harris,A.L., Gearing,A. and Simmons,D.L. (1994) Normal human tissues, in addition to some tumors, express multiple different CD44 isoforms. *Cancer Res.*, **54**, 4539–4546.

Fu,X.-D. (1993) Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature*, **365**, 82–85.

Fu,X.-D. and Maniatis,T. (1992a) The 35-kDa mammalian splicing factor SC35 mediates specific interactions between U1 and U2 small nuclear ribonucleoprotein particles at the 3′ splice site. *Proc. Natl Acad. Sci. USA*, **89**, 1725–1729.

Fu,X.-D. and Maniatis,T. (1992b) Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science*, **256**, 535–538.

Fu,X.-D., Mayeda,A., Maniatis,T. and Krainer,A.R. (1992) General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5′ and 3′ splice site selection. *Proc. Natl Acad. Sci. USA*, **89**, 11224–11228.

Ge,H., Zuo,P. and Manley,J.L. (1991) Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell*, **66**, 373–382.

Gui,J.F., Lane,W.S. and Fu,X.-D. (1994) A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature*, **369**, 678–682.

Harper,J.E. and Manley,J.L. (1992) Multiple activities of the human splicing factor ASF. *Gene Expr.*, **2**, 19–29.

Jackson,D.G. *et al.* (1995) Structure of the CD44 gene and its expression during lymphocyte activation. In Schlossman,S.F. (ed.), *5th International Human Leukocyte Differentiation Antigen Workshop Handbook*. Oxford University Press, Oxford, UK, pp. 1727–1729.

Jiménez-García,L.F. and Spector,D.L. (1993) *In vivo* evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell*, **73**, 47–59.

Kenan,D.J., Query,C.C. and Keene,J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**, 214–220.

Kim,Y.-G., Zuo,P., Manley,J.L. and Baker,B.S. (1992) The *Drosophila* RNA-binding protein RBP1 is localised to transcriptionally active sites of chromosomes and shows a functional similarity to human splicing factor ASF/SF2. *Genes Dev.*, **6**, 2569–2579.

Kohtz,J.D., Jamison,S.F., Will,C.L., Zuo,P., Lührmann,R., García-Blanco,M.A. and Manley,J.L. (1994) Protein–protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**, 119–124.

Krainer,A.R., Maniatis,T., Ruskin,B. and Green,M.R. (1984) Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. *Cell*, **36**, 993–1005.

Krainer,A.R., Conway,G.C. and Kozak,D. (1990) The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. *Genes Dev.*, **4**, 1158–1171.

Krainer,A.R., Mayeda,A., Kozak,D. and Binns,G. (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1-70K, and *Drosophila* splicing regulators. *Cell*, **66**, 383–394.

Lamm,G.M. and Lamond,A.I. (1993) Non-snRNP protein splicing factors. *Biochim. Biophys. Acta*, **1173**, 247–265.

Lavigueur,A., La Branche,H., Kornblihtt,A.R. and Chabot,B. (1993) A splicing enhancer in the human fibronectin ED1 exon interacts with SR proteins and stimulates snRNP binding. *Genes Dev.*, **7**, 2405–2417.

Mattox,W., Ryner,L. and Baker,B.S. (1992) Autoregulation and multifunctionality among *trans*-acting factors that regulate alternative pre-mRNA processing. *J. Biol. Chem.*, **267**, 19023–19026.

Mayeda,A. and Krainer,A.R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell*, **68**, 365–375.

Mayeda,A., Zahler,A.M., Krainer,A.R. and Roth,M.B. (1992) Two members of a conserved family of nuclear phosphoproteins are involved in pre-mRNA splicing. *Proc. Natl Acad. Sci. USA*, **89**, 1301–1304.

Mayeda,A., Helfman,D.M. and Krainer,A.R. (1993) Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol. Cell. Biol.*, **13**, 2993–3001.

Mayeda,A., Munroe,S.H., Cáceres,J.F. and Krainer,A.R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.*, **13**, 5483–5495.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.*, **18**, 7035–7056.

Moore, M.J., Query, C.C. and Sharp, P.A. (1993) Splicing of precursors to mRNAs by the spliceosome. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.

Ring,H.Z. and Lis,J.T. (1994) The SR protein B52/SRp55 is essential for *Drosophila* development. *Mol. Cell. Biol.*, **14**, 7499–7506.

Roth,M.B., Zahler,A.M. and Stolk,J.A. (1991) A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J. Cell Biol.*, **115**, 587–596.

Schmidt,R.E. (1989) Non-lineage/natural killer cell report: new and previously defined clusters. In Knapp,W., Dörkin,B., Gilks,W.R., Rieber,E.P., Schmidt,R.E., Stein,H. and Von dem Bourne,A.E.G. (eds), *Leukocyte Typing IV*. Oxford University Press, Oxford, UK, pp. 533–536.

Staknis,D. and Reed,R. (1994) SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.*, **14**, 7670–7682.

Stephens,C. and Harlow,E. (1987) Differential splicing yields novel adenovirus 5 E1A mRNAs that encode 30 kd and 35 kd proteins. *EMBO J.*, **7**, 2027–2035.

Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.

Sun,Q., Mayeda,A., Hampson,R.K., Krainer,A.R. and Rottman,F.M. (1993) General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes Dev.*, **7**, 2598–2608.

Sureau,A. and Perbal,B. (1994) The PR264/c-myb connection: expression of a splicing factor modulated by a nuclear protooncogene. *Proc. Natl Acad. Sci. USA*, **91**, 932–936.

Tacke,R., Boned,A. and Goridis,C. (1992) ASF alternative transcripts are highly conserved between mouse and man. *Nucleic Acids Res.*, **20**, 5482.

Tanaka,M. and Herr,W. (1990) Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell*, **60**, 375–386.

Tian,M. and Maniatis,T. (1993) A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell*, **74**, 105–114.

Treisman,R., Orkin,S.H. and Maniatis,T. (1983) Specific transcription and RNA splicing defects in five cloned β-thalassaemia genes. *Nature*, **302**, 591–596.

Vellard,M., Sureau,A., Soret,J., Martinerie,C. and Perbal,B. (1992) A potential splicing factor is encoded by the opposite strand of the *trans*-spliced c-myb exon. *Proc. Natl Acad. Sci. USA*, **89**, 2511–2515.

Wu,J.Y. and Maniatis,T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**, 1061–1070.

Zahler,A.M., Lane,W.S., Stolk,J.A. and Roth,M.B. (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.*, **6**, 837–847.

Zahler,A.M., Neugebauer,K.M., Lane,W.S. and Roth,M.B. (1993a) Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science*, **260**, 219–222.

Zahler,A.M., Neugebauer,K.M., Stolk,J.A. and Roth,M.B. (1993b) Human SR proteins and isolation of a cDNA encoding SRp75. *Mol. Cell. Biol.*, **13**, 4023–4028.

Zerler,B., Moran,B., Maruyama,K., Moomaw,J., Grodzicker,T. and Ruley,H.E. (1986) Adenovirus E1A coding sequences that enable ras and pmt oncogenes to transform cultured primary cells. *Mol. Cell. Biol.*, **6**, 887–899.

Zuo,P. and Manley,J.L. (1993) Functional domains of human splicing factor ASF/SF2. *EMBO J.*, **12**, 4727–4737.

Zuo,P. and Manley,J.L. (1994) The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5′ splice sites. *Proc. Natl Acad. Sci. USA*, **91**, 3363–3367.

Received on March 6, 1995; revised on June 27, 1995